ORIGINAL ARTICLE

Personalized surgical recommendations and quantitative therapeutic insights for patients with metastatic breast cancer: Insights from deep learning

Enzhao Zhu[1] | Linmei Zhang[2] | Jiayi Wang[1] | Chunyu Hu[3] | Qi Jing[4] | Weizhong Shi[5] | Ziqin Xu[6] | Pu Ai[1] | Zhihao Dai[7] | Dan Shan[8] | Zisheng Ai[9,10]

[1]School of Medicine, Tongji University, Shanghai, China

[2]Shanghai Engineering Research Center of Tooth Restoration and Regeneration, Research Institute of Stomatology, Stomatological Hospital and Dental School of Tongji University, Shanghai, China

[3]Tenth People's Hospital of Tongji University, School of Medicine, Tongji University, Shanghai, China

[4]Department of Anesthesiology and Perioperative Medicine, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai, China

[5]Shanghai Hospital Development Center, Shanghai, China

[6]Columbia University, New York, NY, USA

[7]School of Medicine, Royal College of Surgeons in Ireland, University of Medicine and Health Sciences, Dublin, Ireland

[8]Department of Biobehavioral Sciences, Columbia University, New York, NY, USA

[9]Department of Medical Statistics, School of Medicine, Tongji University, Shanghai, China

[10]Clinical Research Center for Mental Disorders, Chinese-German Institute of Mental Health, Shanghai Pudong New Area Mental Health Center, School of Medicine, Tongji University, Shanghai, China

Correspondence
Dan Shan, Department of Biobehavioral Sciences, Columbia University, New York, NY 10027, USA.
Email: ds3806@tc.columbia.edu

Zisheng Ai, Department of Medical Statistics, School of Medicine, Tongji University, Shanghai 200331, China.
Email: azs1966@126.com

Abstract

Background: The role of surgery in metastatic breast cancer (MBC) is currently controversial. Several novel statistical and deep learning (DL) methods promise to infer the suitability of surgery at the individual level.

Objective: The objective of this study was to identify the most applicable DL model for determining patients with MBC who could benefit from surgery and the type of surgery required.

Methods: We introduced the deep survival regression with mixture effects (DSME), a semi-parametric DL model integrating three causal inference methods. Six models were trained to make individualized treatment recommendations. Patients who received treatments in line with the DL models' recommendations

Abbreviations: BCS, breast-conserving surgery; DL, deep learning; DRMST, the difference in restricted mean survival time; DSaT, difference in survival probability; DSME, deep survival regression with mixture effects; ESMO, European Society for Medical Oncology; HR, hazard ratio; IBS, integrated Brier score; IPTW, inverse probability weighting; ITE, individual treatment effect; MBC, metastatic breast cancer; NDE, natural direct effect; NIE, natural indirect effect; RD, risk difference.

Enzhao Zhu and Linmei Zhang contributed equally to this study and shared the first authorship.

Dan Shan and Zisheng Ai contributed equally and are corresponding authors.

were compared with those who underwent treatments divergent from the recommendations. Inverse probability weighting (IPW) was used to minimize bias. The effects of various features on surgery selection were visualized and quantified using multivariate linear regression and causal inference.

Results: In total, 5269 female patients with MBC were included. DSME was an independent protective factor, outperforming other models in recommending surgery (IPW-adjusted hazard ratio [HR] = 0.39, 95% confidence interval [CI]: 0.19–0.78) and type of surgery (IPW-adjusted HR = 0.66, 95% CI: 0.48–0.93). DSME was superior to other models and traditional guidelines, suggesting a higher proportion of patients benefiting from surgery, especially breast-conserving surgery. The debiased effect of patient characteristics, including age, tumor size, metastatic sites, lymph node status, and breast cancer subtypes, on surgery decision was also quantified.

Conclusions: Our findings suggested that DSME could effectively identify patients with MBC likely to benefit from surgery and the specific type of surgery needed. This method can facilitate the development of efficient, reliable treatment recommendation systems and provide quantifiable evidence for decision-making.

KEYWORDS

breast surgery, causal inference, deep learning, metastatic breast cancer

1 | INTRODUCTION

Metastatic breast cancer (MBC), often largely incurable, leads to a challenging prognosis for patients, with median survival typically ranging from 2 to 3 years [1, 2]. Currently, the treatment strategies for MBC focus on palliation, and surgical interventions are rare [3–5]. The National Comprehensive Cancer Network (NCCN) Guidelines for Breast Cancer primarily suggest surgery for palliative purposes to address imminent complications or alleviate local symptoms [6]. This approach stems from the prevailing belief that tumor removal may inadvertently accelerate metastatic growth rather than confer survival benefits [7, 8]. However, this perspective has been challenged by several retrospective studies suggesting that surgical removal of primary tumors may in fact enhance survival in patients with MBC [9, 10]. The rationale lies in the potential reduction of circulating tumor cells, including cancer stem cells, and overall tumor burden, which could lead to improved outcomes [11, 12]. For specific groups of patients, such as those with bone-only metastasis or hormone receptor–positive or human epidermal growth factor receptor-2 (HER)-negative tumors, surgery might offer some survival benefits [13, 14], as indicated by the European Society for Medical Oncology (ESMO) Clinical Practice Guidelines [15].

For patients with MBC opting for surgery, mastectomy and breast-conserving surgery (BCS) are the main choices [16]. The trend toward mastectomy is increasingly observed,

often driven by patients' fears of cancer recurrence [17]. Some researchers advocate for BCS wherever feasible, citing benefits including fewer complications and quicker recovery [18]. Nonetheless, the superiority of one surgical approach over another for MBC has not been definitively established through randomized clinical trials. Thus, the selection of surgical methods for patients with MBC remains an area of ongoing investigation. Additionally, advancements in imaging techniques, enabling the detection of smaller metastases, underscore the need for more precise treatment recommendations [19], taking into account the broader implications for the management of primary tumors [20].

This population-based study aimed to identify the most suitable individualized causal inference model for patients with MBC, offering personalized surgical recommendations and determining the appropriate type of surgery. Furthermore, we explored how DL models can discern complex correlations between an individual's characteristics and the potential benefits of various treatments.

2 | METHODS

2.1 | Study design and data source

This retrospective cohort study evaluated the efficacy of DL models in determining the individual treatment effect (ITE) for patients with MBC. We compared systemic treatments

alone versus systemic treatments combined with surgery and also assessed BCS against mastectomy in surgical cases. The surgeries examined included BCS, nipple-sparing mastectomy (NSM), modified radical mastectomy, and radical mastectomy.

Data were sourced from the Surveillance, Epidemiology, and End Results 18 (SEER) database, which covers 18 regions in the United States, representing 27.8% of the population [21]. Our study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [22].

2.2 | Study population and eligibility criteria

This study included female patients initially diagnosed with metastatic ductal, lobular, or ductal-lobular carcinoma between January 1, 2010, and December 31, 2015, and treated with systemic therapy. Cases were excluded if the following criteria were met: (1) unknown or ambiguous demographic information; (2) undefined breast cancer subtypes; (3) unknown TNM stage or tumor size; (4) unspecified metastatic sites; (5) indeterminate histological grades or types; (6) male patients; (7) undetermined surgery types; (8) absence of systemic treatment; (9) bilateral or indeterminate laterality; and (10) incomplete follow-up or multiple malignancies. The cohort selection process is illustrated in Figure 1a.

Tumor stages were determined using the 6th edition of the American Joint Committee on Cancer Staging Manual. All treatment information and other patient characteristics included were recorded at the time of the patient's initial diagnosis. Various curative surgeries were also performed during initial treatment, including SEER code 20: partial mastectomy (BCS), 30: subcutaneous mastectomy (NSM), 40: total mastectomy, 50: modified radical mastectomy, and 60: radical mastectomy. Patients alive as of December 31, 2019, were censored, resulting in a follow-up period of 4–10 years.

2.3 | Calculation of individual treatment effect

In real-world scenarios, only one treatment outcome is observable per patient, with the alternative remaining counterfactual. This counterfactual outcome needs to be predicted.

We defined the outcome as the time it took for the patient to reach 50% mortality, termed time at risk (TaR) [23]. ITE was calculated using Equation (1), where i denotes an individual patient and T denotes different treatments of interest. The ITE reflects the relative efficacy differences for each patient.

The calculation of individual treatment effect is shown in the equation as follows:

$$\text{ITE}_i = TaR_i^{T=1} - TaR_i^{T=0}. \quad (1)$$

2.4 | Deep survival regression with mixture effects and related works

The T-learner adopts two models to estimate ITE as Equation (2), where μ_1 and μ_0 denote models trained on respective treatment groups [24]. Although T-learner mitigates some confounding artifacts, it remains susceptible to inconsistent predictive performance [25] and biased treatment allocation [26].

The individual treatment effect estimation method of T-learner is shown in the equation as follows:

$$\text{ITE} = \mu_1(x) - \mu_0(x). \quad (2)$$

The Balanced Individual Treatment Effect for Survival (BITES) data [26] address these issues using representation-based causal inference. Balancing the generating distributions of treatment groups has been proven to be effective for both covariate space [27] and latent representations [28]. Even within the same treatment group, patients still have variations in risk, limiting the proportional hazard assumption and increasing the likelihood that confounders will be present.

In this research, we introduce the deep survival regression with mixture effects (DSME), a semi-parametric deep learning survival regression model that synthesizes T-learner, representation-based, and subclassification causal inference methods. The architecture of DSME is presented in Figure 1b. DSME contains a shared network and two risk-specific networks. The shared network achieves balanced distributions for generating data by maximizing the p-Wasserstein distance between treatment groups using integral probability metrics (IPM) and calculating smoothed optimal transport loss [29]. Within the risk networks, DSME employs a finite mixture of K neural networks, with the assignment of an individual i to each latent group mediated by a gating function $g(.)$ [30], facilitating the encoding of covariate representations x_i. DSME then takes each minibatch output from the shared network as posterior and maximizes the representations of patients with different risks separately by the $Q(\cdot)$ function [30], in which Monte Carlo expectation maximization is used to weigh the mixture weights and learn the parameters. Because the priori log hazard ratios (HRs) of patients in each latent group are similar and maximized, the heterogeneity in each mixture is subclassified [25]. The risk networks mimic the
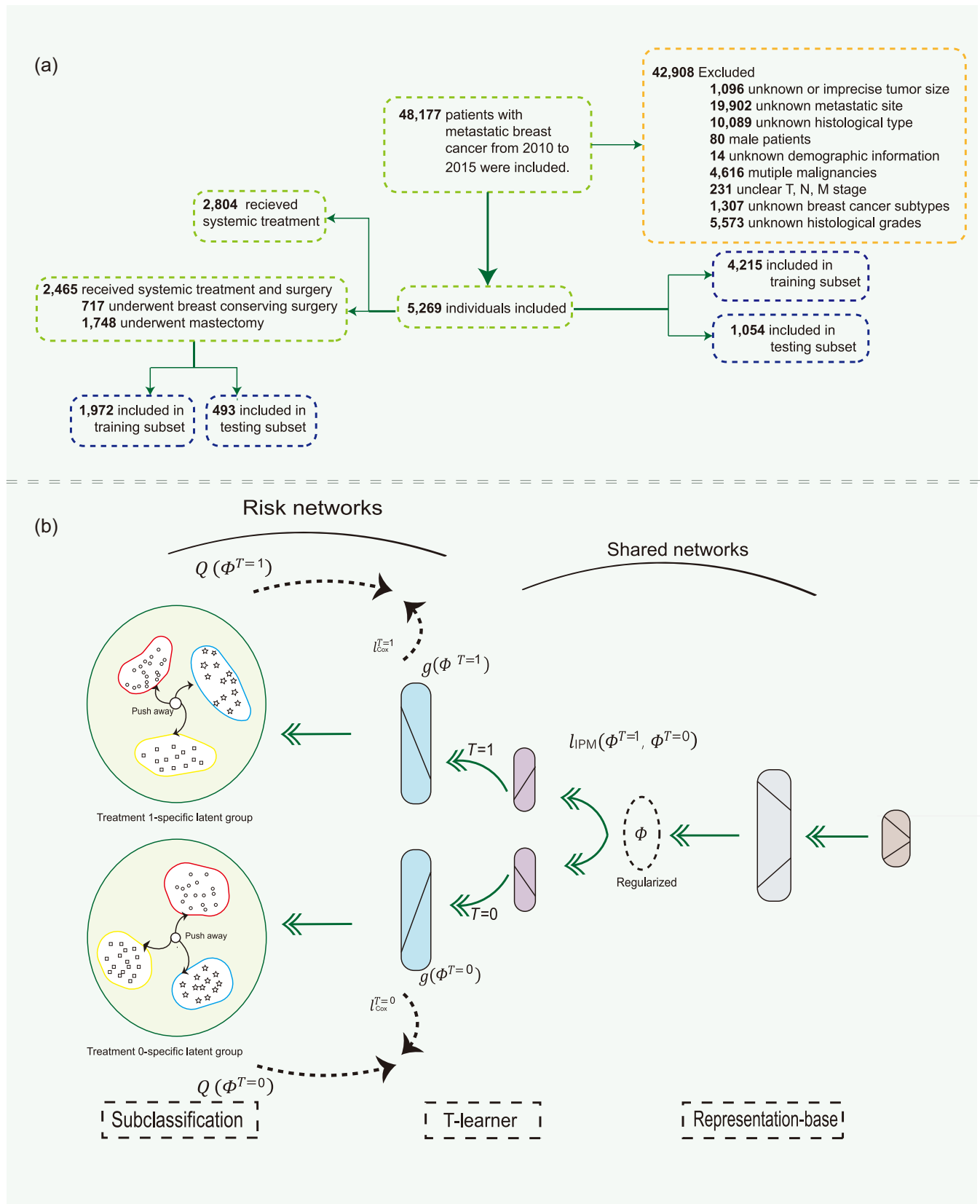
(a)

48,177 patients with metastatic breast cancer from 2010 to 2015 were included.

42,908 Excluded
1,096 unknown or imprecise tumor size
19,902 unknown metastatic site
10,089 unknown histological type
80 male patients
14 unknown demographic information
4,616 mutiple malignancies
231 unclear T, N, M stage
1,307 unknown breast cancer subtypes
5,573 unknown histological grades

2,804 recieved systemic treatment

2,465 received systemic treatment and surgery
717 underwent breast conserving surgery
1,748 underwent mastectomy

5,269 individuals included

4,215 included in training subset

1,054 included in testing subset

1,972 included in training subset

493 included in testing subset

(b)

FIGURE 1 Diagram of the inclusion procedure and model architecture. (a) Patients inclusion flowchart. (b) The architecture of deep survival regression with mixture effects.

T-leaner architecture, with each handling data from their respective treatment cohort. The Cox partial log-likelihood is independently computed in different risk networks. The loss function of DSME is formulated as Equation (3), where q is the fraction of patients in the treatment 0 cohort, E denotes a failure event, y denotes observed survival times, and l_{Cox} is the negative Cox partial log-likelihood. The overall strength of IPM regularization is adjusted by α, and the hazard function $h(\cdot)$ is deduced from the regularized latent representation Φ. The algorithms for l_{IPM} [26], l_{Cox} [26], and $Q(\cdot)$ functions [30] are in line with previous studies. DSME computes treatment-specific baseline hazards at inference.

The loss function of DSME is shown in the equation as follows:

$$
\begin{aligned}
l_{DSME}(x_i, y_i, E_i, T_i) = {} & q l_{Cox}^{T=0}(h_0(\Phi(x)), Y^{T=0}, E^{T=0}) \\
& + (1-q) l_{Cox}^{T=1}(h_1(\Phi(x)), Y^{T=1}, E^{T=1}) \\
& + \alpha l_{IPM_\epsilon^p}(\Phi^{T=1}, \Phi^{T=0}) \\
& + Q^{T=0}(\Phi^{T=0}) + Q^{T=1}(\Phi^{T=1})
\end{aligned}
\tag{3}
$$

2.5 | Model development and validation

All patients were randomly assigned to either the training set (consisting of 80% of the samples) for model training or the testing set (comprising 20% of the samples) for assessing the models' performance. During the training period, we used fivefold cross-validation to tune the models' hyperparameters, while the testing set remained hidden. To maintain consistency across patient data, the same cohort of patients was used in both the training and testing sets for the surgery type recommendation models as was used for the initial surgery recommendation models, with the sole exception being the exclusion of patients who did not undergo any surgery. Each model was trained independently for the two phases of recommendation tasks while sharing the same architectural framework.

Treatment recommendations of the model can be obtained using the value of ITE. To explore the protective effect of the model's recommendation, we divided patients into the recommended (Consis.) and anti-recommended (Inconsis.) groups, on the basis of the actual treatment's congruence with model advice.

The Cox proportional hazards model (CPH) and random survival forest (RSF) are commonly used machine learning models for survival prediction. DeepSurv [31] contains a core hierarchical structure comprising fully connected feed-forward neural networks (five-layered) and a single output node. This model calculates patient survival risks via the negative log-partial likelihood function. Contrary to CPH, DeepSurv relaxes the assumptions of

data normality and variance alignment. These three models mentioned above were implemented as part of a T-learner framework, where two base models are independently developed across different treatment groups.

The Cox Mixtures with Heterogeneous Effects (CMHE) model [32] uses a latent variable approach to capture heterogeneous treatment effects. It presupposes that individuals may fall into one of several latent clusters, each characterized by unique response patterns. The assignment function of an individual to a latent group also allows the model to learn jointly with the component-specific HRs that relax the proportional hazard assumption.

Additionally, we trained and evaluated DSME and BITES as comparisons. Altogether, six models were developed and subjected to evaluation.

2.6 | Statistical analysis

Data analysis was performed using R version 4.1.3 and Python 3.8. Continuous variables are reported as medians and interquartile ranges (IQRs), and categorical variables are presented as numbers and percentages (%). Inverse probability weighting (IPW) and propensity score matching (PSM) were applied to mitigate confounding and bias in treatment selection. A marginal structural cause-specific Cox proportional hazards model (MSM) [33] was used to analyze competing risks. The Log-rank test was used to compare the Kaplan–Meier (KM) survival curves.

3 | RESULTS

3.1 | Patients

In total, 5269 female patients with MBC with a median follow-up time of 37 months (IQR: 16–61 months) were included. The median interval from diagnosis to the initiation of treatment was 1 month (IQR: 0–1 month). Of the total patients, 2465 (46.8%) underwent breast surgery, with 717 (29.1%) receiving BCS and 1748 (70.9%) that underwent mastectomy. The overall mortality rate was 70.8% (95% confidence interval [CI]: 69.5%–72.0%). The demographic and clinical characteristics of each treatment group are presented in Table 1.

3.2 | Performance

In our analysis, we had a total of 5269 patients for surgery recommendation and 2465 patients for surgery type recommendation, after manually excluding those who did not undergo surgery. The testing set for performance evaluation comprised 1054 patients for

TABLE 1 Study population.

Characteristics	Systemic treatment (n = 2804)	Systemic treatment plus surgery (n = 2465)	Breast-conserving surgery (n = 717)	Mastectomy (n = 1748)
Age, median (IQR), years	58.0 (49.0–66.0)	57.0 (47.0–66.0)	58.0 (49.0–66.0)	56.0 (46.0–66.0)
Tumor size, median (IQR), mm	42.0 (27.0–65.0)	60.0 (27.0–68.0)	30.0 (22.0–42.0)	50.0 (30.0–75.0)
Married (%)	1296 (46.2)	1225 (49.7)	381 (53.1)	844 (48.3)
Race—White (%)	2066 (73.7)	1849 (75.0)	559 (78.0)	1290 (73.8)
Income—higher than 70,000$ (%)	939 (33.5)	786 (31.9)	251 (35.0)	535 (30.6)
Grade (%)				
I	191 (6.8)	136 (5.5)	51 (7.1)	85 (4.9)
II	1252 (44.7)	897 (36.4)	269 (37.5)	628 (35.9)
III	1342 (47.9)	1420 (57.6)	395 (55.1)	1025 (58.6)
IV	19 (0.7)	12 (4.9)	2 (0.3)	10 (0.6)
Location (%)				
Upper outer quadrant	799 (28.5)	695 (28.2)	223 (31.1)	472 (27.0)
Upper inner quadrant	210 (7.5)	181 (7.3)	89 (12.4)	92 (5.3)
Lower outer quadrant	177 (6.3)	167 (6.8)	66 (9.2)	101 (5.8)
Lower inner quadrant	125 (4.5)	110 (4.5)	50 (7.0)	60 (3.4)
Central/overlapping	774 (27.6)	723 (29.3)	176 (24.5)	547 (31.3)
Nipple/axillary tail	27 (1.0)	22 (0.9)	7 (1.0)	5 (0.9)
Other/unknown	692 (24.7)	567 (23.0)	106 (14.8)	461 (26.4)
T stage (%)				
T1	367 (13.1)	278 (11.3)	148 (20.6)	130 (7.4)
T2	963 (34.3)	1027 (41.7)	394 (55.0)	633 (36.2)
T3	572 (20.4)	491 (19.9)	80 (11.2)	411 (23.5)
T4	902 (32.2)	669 (27.1)	95 (13.2)	547 (32.8)
N stage (%)				
N0	527 (18.8)	354 (14.4)	193 (26.9)	161 (9.2)
N1	1223 (43.6)	826 (33.5)	247 (34.4)	579 (33.1)
N2	182 (6.5)	486 (19.7)	110 (15.3)	376 (21.5)
N3	872 (31.1)	799 (32.4)	167 (23.3)	632 (36.2)
Distant metastasis (%)				
Bone	1989 (70.9)	1600 (64.9)	471 (65.7)	1129 (64.6)
Brain	234 (8.3)	98 (4.0)	35 (4.9)	63 (3.6)
Liver	958 (34.2)	559 (22.7)	161 (22.5)	398 (22.8)
Lung	993 (35.4)	618 (25.1)	149 (20.8)	469 (26.8)
ER status—positive	2087 (74.1)	1805 (73.2)	537 (74.9)	1268 (72.5)
PR status—positive	1687 (60.2)	1442 (58.5)	432 (60.3)	1010 (57.8)
Subtypes (%)				
HR+/HER2−	1523 (54.3)	1376 (55.8)	414 (57.7)	962 (55.0)
HR−/HER2−	380 (13.6)	378 (15.3)	103 (14.4)	275 (15.7)
HR+/HER2+	603 (21.5)	461 (18.7)	131 (18.3)	330 (18.9)
HR−/HER2+	298 (10.6)	250 (10.1)	69 (9.6)	181 (10.4)

TABLE 1 (Continued)

Characteristics	Systemic treatment (*n* = 2804)	Systemic treatment plus surgery (*n* = 2465)	Breast-conserving surgery (*n* = 717)	Mastectomy (*n* = 1748)
Axillary lymph node status (%)				
Positive	1000 (35.7)	1732 (70.3)	391 (54.5)	1341 (76.7)
Negative	40 (1.4)	296 (12.0)	105 (14.6)	191 (10.9)
Not evaluated	1764 (62.9)	437 (17.7)	221 (30.8)	216 (12.4)
Regional lymph node status (%)				
Positive	1059 (37.8)	1765 (71.6)	397 (55.4)	1368 (78.3)
Negative	53 (1.9)	324 (13.1)	111 (15.5)	213 (12.2)
Not evaluated	1683 (60.0)	371 (15.1)	207 (28.9)	163 (9.4)
Adjuvant treatment (%)				
Radiotherapy	875 (31.2)	1284 (52.1)	413 (57.6)	871 (49.8)
Chemotherapy	2223 (79.3)	1858 (75.4)	500 (69.7)	1358 (77.7)
Follow-up, median (IQR), months	31 (12–53)	47 (21–69)	52.0 (24.0–76.0)	44.0 (19.8–67.0)

Abbreviations: ER, estrogen receptor; IQR, interquartile range; PR, progesterone receptor.

surgery recommendation and 493 for surgery type recommendation; the results are detailed in Table 2.

To assess model discrimination, we calculated the integrated Brier score (IBS) for each treatment group, which is acknowledged as a measure of phenotyping purity [33]. In the surgery recommendation task, CPH had the best discrimination (IBS in systemic treatment group (IBS^b): 0.17, 95% CI: 0.15–0.18; IBS in systemic treatment plus surgery group (IBS^c): 0.18, 95% CI: 0.16–0.19), with RSF closely following (IBS^b: 0.17, 95% CI: 0.16–0.18; IBS^c: 0.18, 95% CI: 0.16–0.19). Both CPH and RSF also showed robust performance in the surgery type recommendation task.

To investigate the protective effect of adherence to model recommendations, we calculated various metrics: HR, difference in restricted mean survival time within 5 years (DRMST), difference in survival probability at 5 years (DSaT), and risk difference (RD). We used IPW to adjust for all covariates and actual treatment between Consis. and Inconsis. groups, therefore enhancing objectivity and reducing bias. We also compared the 2021 ESMO guidelines for MBC; patients whose actual treatment was consistent with the guidelines were categorized in the Consis. group and the rest were in the Inconsis. group.

DSME performed the best in surgery recommendation (HR = 0.63, 95% CI: 0.54–0.74; IPW-adjusted HR (HR^a) = 0.39, 95% CI: 0.19–0.78; DRMST = 7.06, 95% CI: 4.46–9.66; IPW-adjusted DRMST ($DRMST^a$) = 7.85, 95% CI: 5.01–10.23; DSaT = 31.67%, 95% CI: 15.21%–46.88%; IPW-adjusted DSaT ($DSaT^a$) = 33.80%, 95% CI: 19.00%–59.19%; RD = 15.20%, 95% CI: 5.30%–17.90%; IPW-adjusted RD (RD^a) = 3.70%, 95% CI: 7.68%–19.70%), outperforming the ESMO guidelines

(HR = 0.67, 95% CI: 0.57–0.78; HR^a = 1.00, 95% CI: 0.58–1.73; DRMST = 5.56, 95% CI: 2.94–8.18; $DRMST^a$ = 5.60, 95% CI: 2.71–7.97; DSaT= 31.41%, 95% CI: 11.01%–49.89%; $DSaT^a$ = 30.66%, 95% CI: 11.01%–49.89%; RD = 14.20%, 95% CI: 7.90%–20.50%; RD^a = 11.10%, 95% CI: 4.47%–17.80%), other models, and simply referring all patients for surgery.

In surgery type recommendations, DSME (HR = 0.77, 95% CI: 0.60–0.99; HR^a = 0.66, 95% CI: 0.48–0.93; DRMST = 4.78, 95% CI: 0.99–8.57; $DRMST^a$ = 4.75, 95% CI: 0.41–8.65; DSaT = 40.13, 95% CI: 8.89%–49.02%; $DSaT^a$ = 39.95%, 95% CI: 1.31%–55.43%; RD = 8.90%, 95% CI: 0.01%–18.70%; RD^a = 5.28%, 95% CI: 3.93%–8.50%) maintained its superior performance, being the only model to consistently show a statistically significant protective effect across both treatment recommendation tasks.

According to DSME's insights, 97.8% of patients could have a survival benefit from receiving breast surgery and 73.6% were more suited for BCS over mastectomy.

The protective effect of DSME because of an imbalance in the surgery proportions in the two groups was also of interest. Thus, we treated surgery as a mediator and adjusted for all baseline features to calculate the natural direct effect (NDE) and natural indirect effect (NIE), detailed in Figure 2a. Similarly, the type of surgery was treated as a mediator in the evaluation of the second stage (Figure 2b). These values are presented as the slope of a linear regression. NDE measures the direct effect of DSME on mortality reduction, excluding the effect of actual treatment. The DSME recommendation was statistically significant in both the surgery (NDE: −0.12, 95% CI: −0.15

CANCER INNOVATION

TABLE 2 Detailed model performance and treatment recommendation effect.

Systemic treatment versus systemic treatment plus surgery

Model	HR	HR[a]	DRMST	DRMST[a]	DSaT (%)	DSaT[a] (%)	RD (%)	RD[a] (%)	IBS[b]	IBS[c]
DSME ($n^g = 23$)	**0.63 (0.54–0.74)**	**0.39 (0.19–0.78)**	**7.06 (4.46–9.66)**	**7.85 (5.01–10.23)**	31.67 (15.21–46.88)	**33.80 (19.00–59.19)**	**15.20 (5.30–17.90)**	**13.70 (7.68–19.70)**	0.18 (0.17–0.19)	0.20 (0.19–0.21)
BITES ($n^g = 219$)	0.72 (0.61–0.84)	0.86 (0.70–1.06)	5.59 (2.97–8.21)	5.58 (3.10–8.38)	**33.16 (11.1–45.11)**	33.36 (6.94–48.52)	11.60 (5.30–17.90)	10.00 (4.04–16.10)	0.18 (0.17–0.20)	0.18 (0.17–0.19)
DeepSurv ($n^g = 264$)	0.78 (0.67–0.91)	0.92 (0.76–1.12)	3.42 (0.78–6.06)	3.42 (0.37–6.15)	34.84 (8.02–42.86)	34.66 (3.50–45.77)	8.00 (1.70–14.30)	7.96 (1.92–14.00)	0.26 (0.23–0.29)	0.28 (0.25–0.30)
CMHE ($n^g = 979$)	1.48 (1.26–1.74)	Reference	−5.52 (−8.16 to −2.89)	−6.54 (−11.16 to −3.49)	32.74 (−12.98 to 46.72)	32.50 (−21.26 to 52.45)	−14.00 (−20.30 to −7.70)	−12.10 (−18.20 to −6.02)	0.19 (0.18–0.20)	0.20 (0.19–0.21)
CPH ($n^g = 81$)	0.64 (0.55–0.76)	0.64 (0.25–1.68)	6.65 (4.03–9.27)	6.69 (6.43–6.92)	33.37 (12.99–46.36)	32.11 (9.22–48.75)	13.00 (6.50–19.50)	14.90 (5.65–24.10)	**0.17 (0.15–0.18)**	**0.18 (0.16–0.19)**
RSF ($n^g = 98$)	0.68 (0.58–0.80)	0.81 (0.61–1.07)	5.65 (3.01–8.27)	5.67 (3.26–7.78)	33.48 (11.75–45.23)	32.72 (7.60–48.91)	11.80 (5.50–18.10)	9.97 (3.92–16.00)	0.17 (0.16–0.18)	0.18 (0.16–0.19)
ARS ($n^g = 0$)	0.65 (0.56–0.77)	0.70 (0.57–0.86)	6.09 (3.48–8.70)	4.21 (1.86–6.77)	31.86 (14.42–46.28)	33.58 (3.14–50.12)	14.40 (8.10–20.70)	11.80 (5.57–18.00)	–	–
ESMO ($n^g = 41$)	0.67 (0.57–0.78)	1.00 (0.58–1.73)	5.56 (2.94–8.18)	5.60 (2.71–7.97)	31.41 (11.01–49.89)	30.66 (11.01–49.89)	14.20 (7.90–20.50)	11.10 (4.47–17.80)	–	–

Breast-conserving surgery versus mastectomy

Model	HR	HR[d]	DRMST	DRMST[d]	DSaT (%)	DSaT[d] (%)	RD	RD[d] (%)	IBS[e]	IBS[f]
DSME ($n^h = 130$)	**0.77 (0.60–0.99)**	**0.66 (0.48–0.93)**	**4.78 (0.99–8.57)**	**4.75 (0.41–8.65)**	40.13 (8.89–49.02)	**39.95 (1.31–55.43)**	8.90 (0.01–18.70)	**5.28 (3.93–8.50)**	0.19 (0.18–0.20)	0.19 (0.19–0.20)
BITES ($n^h = 214$)	0.82 (0.65–1.03)	0.90 (0.68–1.20)	3.52 (−0.15–7.18)	3.61 (0.17–5.53)	38.03 (10.07–48.09)	38.03 (2.56–53.99)	**10.10 (0.90–18.70)**	8.01 (−0.98 to 17.00)	0.19 (0.16–0.21)	0.17 (0.15–0.19)
DeepSurv ($n^h = 92$)	0.89 (0.70–1.13)	0.96 (0.67–1.38)	2.01 (−1.77 to 5.79)	1.94 (−3.14 to 7.00)	41.93 (3.42–45.34)	40.97 (−3.96 to 50.98)	3.40 (−6.20 to 13.00)	1.42 (−7.61 to 10.40)	0.25 (0.19–0.31)	0.28 (0.25–0.32)
CMHE ($n^h = 174$)	0.96 (0.75–1.22)	Reference	0.89 (−2.96 to 4.74)	0.48 (−3.69 to 4.43)	43.13 (−0.34 to 43.17)	41.32 (−7.84 to 50.03)	0 (−9.80 to 9.80)	0.60 (−0.86 to 9.84)	0.21 (0.19–0.22)	0.21 (0.20–0.22)
CPH ($n^h = 148$)	0.96 (0.75–1.23)	0.83 (0.62–1.13)	0.39 (−3.49 to 4.28)	0.24 (−3.71 to 4.31)	**43.01 (0.62–43.63)**	40.97 (−7.17 to 49.95)	0.60 (−9.20 to 10.40)	0.45 (−8.69 to 9.59)	**0.16 (0.13–0.18)**	**0.17 (0.15–0.19)**
RSF ($n^h = 131$)	0.86 (0.67–1.1)	0.92 (0.67–1.11)	2.18 (−1.71 to 6.06)	2.27 (−1.50 to 5.54)	41.14 (6.22–47.40)	40.91 (−1.67 to 53.85)	6.20 (−3.80 to 16.20)	3.72 (−5.57 to 13.00)	0.17 (0.15–0.19)	0.18 (0.16–0.19)
ARB ($n^h = 0$)	0.90 (0.69–1.16)	0.91 (0.67–1.23)	1.29 (−2.75 to 5.33)	1.33 (−3.48 to 4.85)	42.88 (1.09–43.97)	41.61 (−8.74 to 51.80)	1.10 (−9.30 to 11.50)	−1.21 (−10.70 to 8.25)	–	–

Note: HR, univariate hazards ratio; HR[a], multivariate hazards ratio that adjusted for all covariates and surgery using inverse probability weighting (IPW); DRMST, the difference in restricted mean survival time within 5 years; DRMST[a], 5-year DRMST adjusted for all covariates and surgery using IPW; DSaT, the difference in survival probability at 5 years; DSaT[a], 5-year DSaT adjusted for all covariates and surgery using IPW; RD, risk difference at 5 years; IBS[b], integrated Brier score in the systemic treatment group; IBS[c], integrated Brier score in the systemic treatment plus surgery group; HR[d], multivariate hazards ratio that adjusted for all covariates and surgery types using IPW; DRMST[d], 5-year DRMST adjusted for all covariates and surgery types using IPW; DSaT[d], 5-year DSaT adjusted for all covariates and surgery types using IPW; IBS[e], integrated Brier score in the breast-conserving surgery (BCS) group; IBS[f], integrated Brier score in the mastectomy group; DSME, Deep Survival regression with Mixture Effects; BITES, Balanced Individual Treatment Effect for Survival data; CMHE, Cox Mixtures with Heterogeneous Effects; CPH, Cox proportional hazards model; RSF, random survival forest; ARS, all recommend for surgery; ARB, all recommend for mastectomy; Reference, model not fitted. Bolded font indicates that the model performs best in this metric. In ESMO guidelines, breast surgery is considered for patients with hormone receptor-positive tumors, human epidermal growth factor receptor-2-negative tumors, patients younger than 55 years of age, and patients who respond well to initial systemic therapy. n^g, number of patients who were recommended for surgery; n^h, number of patients who were recommended for mastectomy; ESMO, European Society for Medical Oncology.
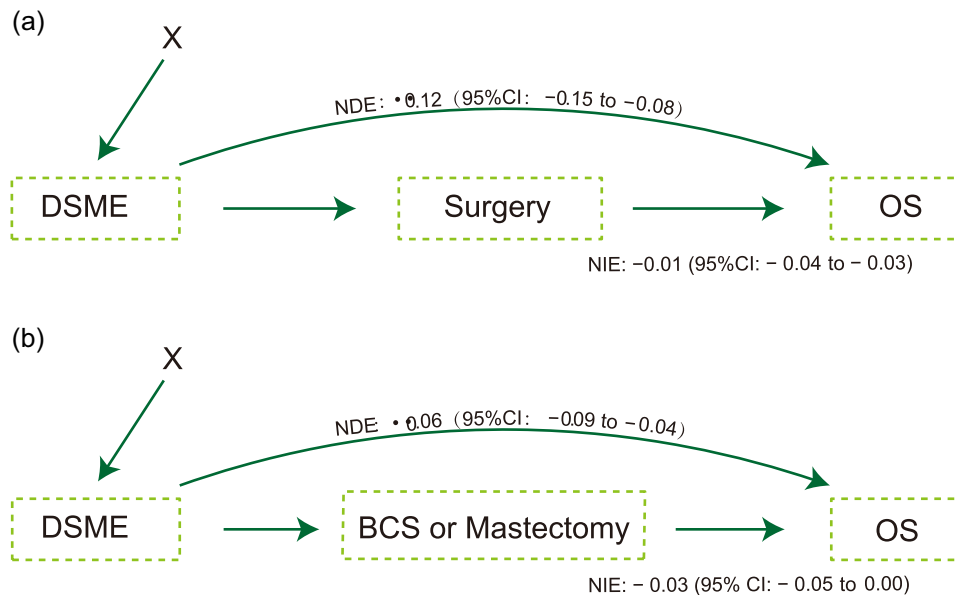
(a)

(b)

FIGURE 2 Causal path of model recommendation. (a) Causal path deep survival regression with mixture effects regarding surgery recommendation. (b) Causal path deep survival regression with mixture effects regarding surgery type recommendation. BCS, breast-conserving surgery; DSME, deep survival regression with mixture effects; NDE, natural direct effect; NIE, natural indirect effect; OS, overall survival; X denotes patient covariates.

to −0.08; NIE: −0.01, 95% CI: −0.04 to −0.03) and the surgery type (NDE: −0.06, 95% CI: −0.09 to −0.04; NIE: −0.03, 95% CI: −0.05 to 0.00) tasks.

We also assessed the protective effect of DSME on various causes of death, as shown in Supporting Information: Table S1. Accounting for competing risks, when a specific cause of death was examined, other causes were considered competing events. The HR with competing risks (HR^d) was calculated using MSM. Adoption of the surgery recommendation led to reduced mortality from breast cancer (HR^d = 0.66, 95% CI: 0.61–0.70, $p < 0.001$; IPW-adjusted HR^d = 0.69, 95% CI: 0.59–0.80, $p < 0.001$), miscellaneous malignant cancer (HR^d = 0.57, 95% CI: 0.33–0.98, $p < 0.001$; IPW-adjusted HR^d = 0.30, 95% CI: 0.04–0.97, $p = 0.032$), and liver diseases (HR^d = 1.16, 95% CI: 0.23–3.22, $p = 0.830$; IPW-adjusted HR^d = 0.42, 95% CI: 0.15–0.97, $p = 0.042$). For surgery type recommendation, breast cancer (HR^d = 0.77, 95% CI: 0.68–0.86, $p < 0.001$; IPW-adjusted HR^d = 0.81, 95% CI: 0.65–0.92, $p = 0.044$) and adverse effect (HR^d = 0.96, 95% CI: 0.30–3.06, $p = 0.940$; IPW-adjusted HR^d = 0.62, 95% CI: 0.29–0.92, $p = 0.021$) mortality also reduced.

3.3 | Average treatment effect and treatment heterogeneity

The average treatment effect of surgery and mastectomy is presented in Figure 3a,b. We used IPW and 1:1 PSM to adjust for age, tumor size, histological grades, N status,

metastatic sites, breast cancer subtypes, tumor locations, laterality, and radiotherapy. For surgery, the standardized mean differences (SMDs) of all patients, patients who met the ESMO guidelines, patients who did not meet the ESMO guidelines, patients recommended for surgery by DSME, and patients not recommended for surgery by DSME are presented in Figure S1a–e. Similarly, for mastectomy compared with BCS, the SMDs of all patients, BCS recommended by DSME, and mastectomy recommended by DSME are illustrated in Figure S2a–c.

The KM curves for overall survival (OS) comparing surgery with non-surgery and BCS with mastectomy are displayed in Figure S3a,b. The surgery group demonstrated significantly improved OS compared with the non-surgery group ($p < 0.0001$; IPW-adjusted $p < 0.0001$). Although mastectomy initially appeared to have a survival advantage over BCS before IPW correction ($p < 0.0001$), this advantage was not maintained after correction (IPW-adjusted $p = 0.5357$).

SMDs indicated that both IPW and PSM effectively balanced prognostic features, with PSM being much more effective in certain groups, including patients who did not meet ESMO guidelines, patients not recommended for surgery by DSME, BCS recommended by DSME, and mastectomy recommended by DSME.

Analysis revealed that surgery generally served as a protective factor across all patient groups. However, the protective benefit was not statistically significant for patients whose DSME did not recommend surgery (HR = 0.72, 95%
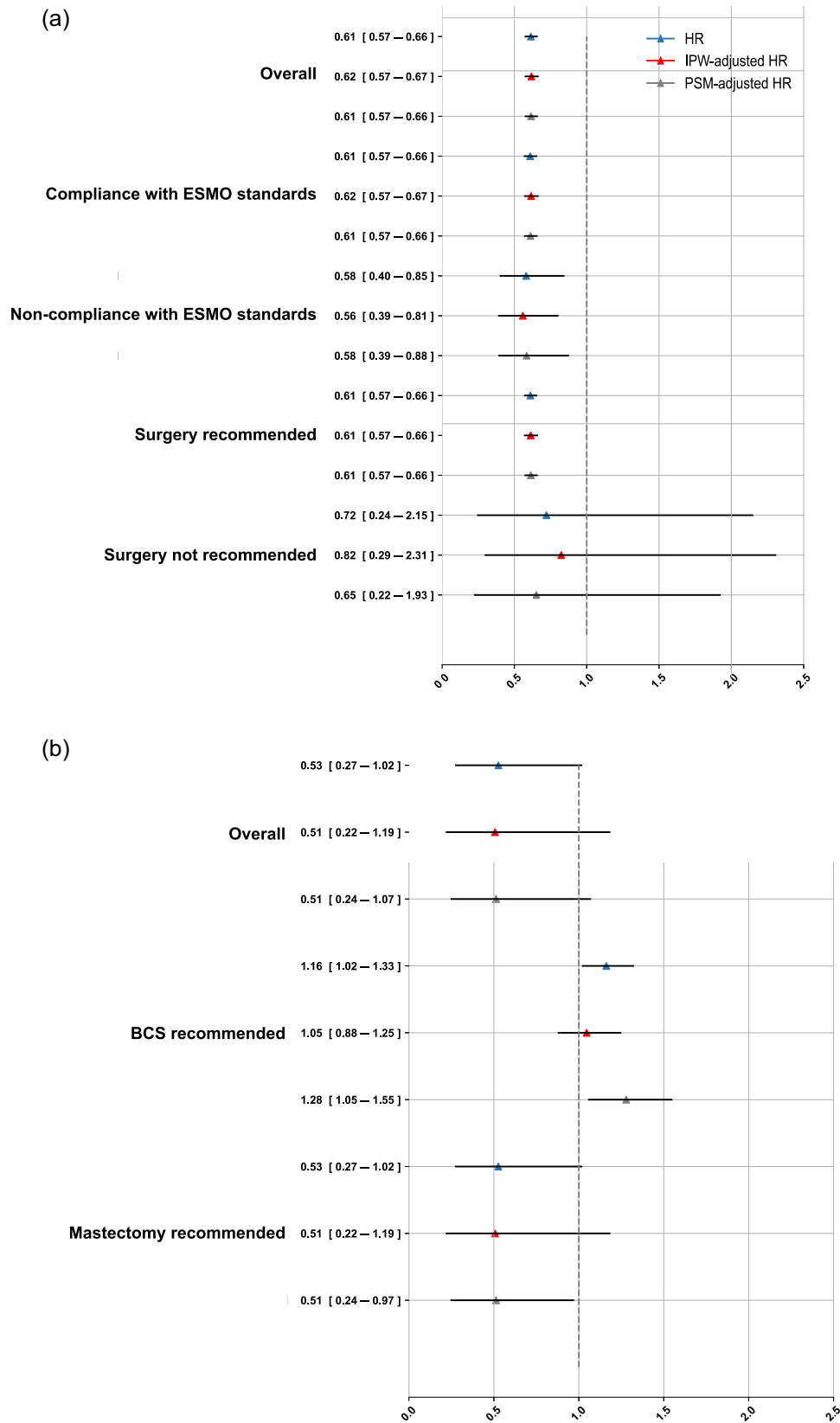
(a)

(b)

FIGURE 3 Average treatment effect and treatment heterogeneity. (a) Average treatment effect of breast surgery. (b) Average treatment effect of mastectomy compared with breast-conserving surgery. BCS, breast-conserving surgery; ESMO, European Society for Medical Oncology; HR, multivariate hazard ratio; IPW, inverse probability weighting; PSM, 1:1 propensity score matching.

CI: 0.24–2.15; IPW-adjusted HR = 0.82, 95% CI: 0.29–2.31; PSM-adjusted HR = 0.65, 95% CI: 0.22–1.93).

Mastectomy did not demonstrate a significant advantage over BCS in the overall patient population. In contrast, within the subgroup recommended for mastectomy by DSME, mastectomy showed a favorable trend (HR = 0.53, 95% CI: 0.27–1.02; IPW-adjusted HR = 0.51, 95% CI: 0.22–1.19; PSM-adjusted HR = 0.51, 95% CI: 0.24–0.97). Conversely, in the subgroup where BCS was the recommended treatment, mastectomy was less beneficial (HR = 1.16, 95% CI: 1.02–1.33; IPW-adjusted HR = 1.05, 95% CI: 0.88–1.25; PSM-adjusted HR = 1.28, 95% CI: 1.05–1.55).

3.4 | Therapeutic insights

A multivariate linear regression model was used to predict ITE from patient covariates. The resulting beta coefficients suggest that the presence or increase of a feature, with all other variables held constant, corresponds to an average change in ITE, indicative of the amount of increase in surgery efficacy. These beta coefficients are depicted in Figure 4a.

For each increase in age (−0.73, 95% CI: −0.77 to −0.69), tumor size (−0.16, 95% CI: −0.17 to −0.15), and metastatic sites (−3.60, 95% CI: −5.20 to −2.01), the time
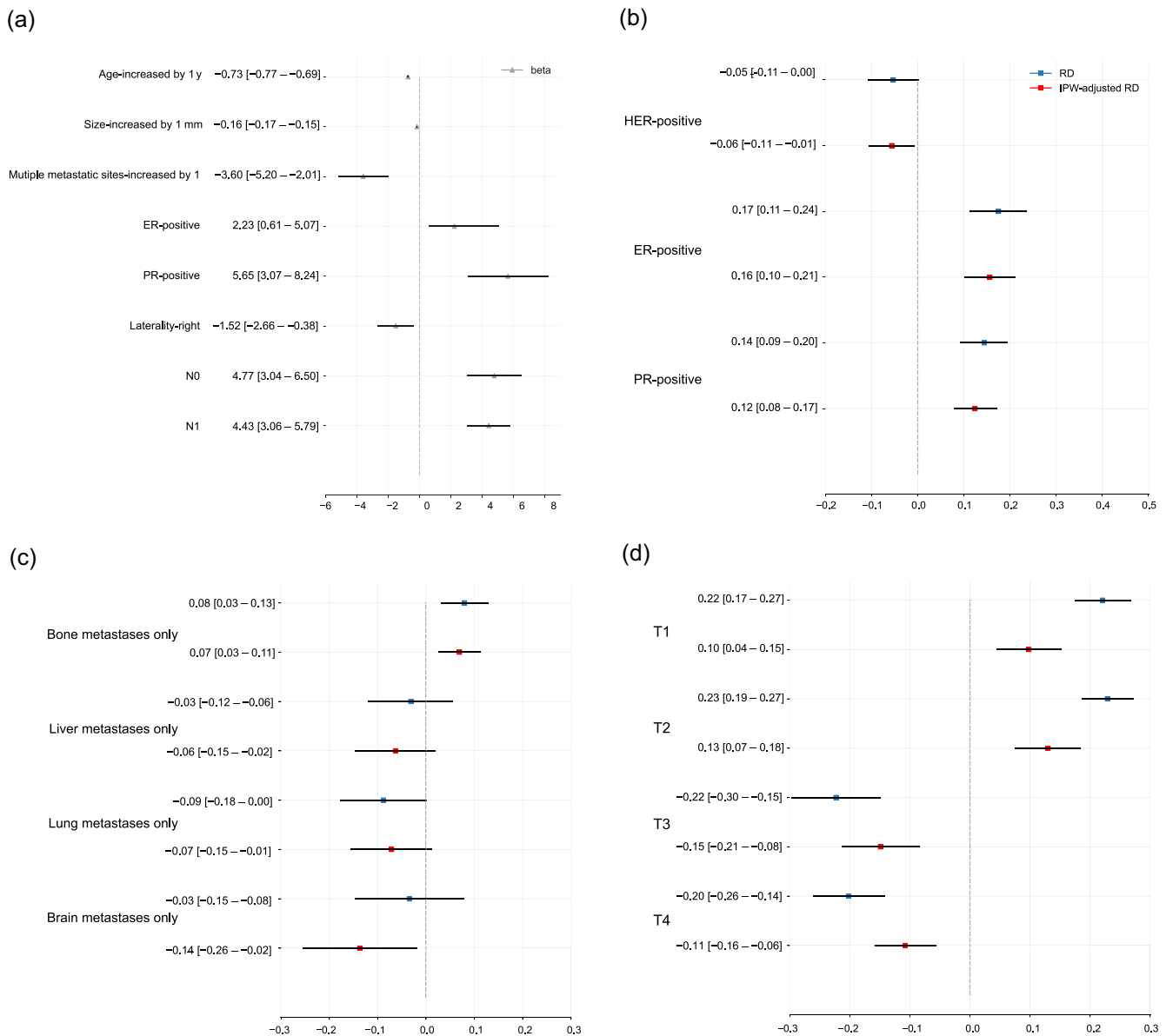

FIGURE 4 Therapeutic insights from deep survival regression with mixture effects. (a) The effect of features on continuous changes in the individual treatment effect. (b) The quantified effect of breast cancer subtypes on surgery selection. (c) The quantified effect of metastatic sites on surgery selection. (d) The quantified effect of tumor size on surgery selection. ER, estrogen receptor; HER, human epidermal growth factor receptor-2; IPW, inverse probability weighting; PR, progesterone receptor; RD, risk difference, indicating the probability that a patient with this feature was recommended for surgery minus the probability under the condition that the patient was without this feature.

to 50% mortality was reduced for patients undergoing surgery compared with those not undergoing surgery. Right laterality (−1.52, 95% CI: −2.66 to −0.40) was also a factor that contributed to reduced surgical effectiveness. Conversely, estrogen receptor (ER) positivity (2.23, 95% CI: 0.61–5.07), progesterone receptor (PR) positivity (5.65, 95% CI: 3.07–8.24), and lower lymph node involvement (N0: 4.77, 95% CI: 3.04–6.50; N1: 4.43, 95% CI: 3.06–5.79) corresponded with improved outcomes.

By considering patient characteristics as key variables, we transformed the analysis of their effect on surgery selection into a binary causal inference problem. Using this methodology, RD and IPW-adjusted RD, which include factors such as breast cancer subtypes, metastatic sites, and tumor size, were computed as shown in Figure 4b–d. The RD represents the difference in the probability of being recommended for surgery attributable to a particular feature, while the IPW-adjusted RD provides a more unbiased estimate by accounting for other variables and their interactions.

Patients with ER positivity, PR positivity, bone-only metastases, and smaller tumors were more likely to be recommended for surgery. In contrast, HER2-positive status, brain-only metastases, and larger tumor size decreased the likelihood of a surgical recommendation.

3.5 | Model interpretation based on SurvSHAP(t)

SurvSHAP(t) [34], a pioneering approach offering time-dependent explanations for survival regression in DL models, was used to elucidate DSME's outputs. Figure 5a,b visualizes the aggregation of the eight most influential variables, sorted by their aggregated Sharpley values and rankings over 300 observations in the testing set, for the surgery and surgery type recommendation models, respectively.

The horizontal bars graphically represent the number of observations for which the importance of the variable, represented as a given color, was ranked as first, second, and so on. The treatment of interest represented the usage of different risk networks and baseline hazards.

4 | DISCUSSION

We introduced the DSME, an innovative approach amalgamating representation-based, T-learner, and subclassification causal inference methods. Upon careful validation and stringent bias mitigation, DSME exhibited efficacy in extending the survival of patients with MBC by 8 months over a 5-year span. This performance surpasses that of real-world decisions by clinicians, contemporary models, ESMO guidelines, and generic treatment approaches focused on average outcomes. Despite similarities in recommendation trends with ESMO guidelines and prior research, DSME showed superior discernment of potential treatment heterogeneity, likely from its intricate handling of complex feature interactions that have been inadequately addressed in the existing literature. Thus, DSME represents a promising tool for clinical decision-making for MBC patient care.

The debate concerning varied surgical interventions for MBC has persisted [14, 18, 35, 36]. Harbeck et al. suggested that patients with MBC should receive more personalized treatment, considering the great heterogeneity within this group [37]. Our findings suggested that almost all patients with MBC could benefit from surgical intervention, independent of ESMO guideline adherence. For patients where DSME does not recommend surgery, the absence of a significant protective effect suggests surgery may introduce unnecessary risks and delay essential systemic therapy, potentially impairing overall outcomes [38, 39].

The dichotomy between BCS and mastectomy reveals treatment heterogeneity, with mastectomy demonstrating a protective effect when recommended and BCS appearing as the favorable option otherwise. These distinctions further underscore DSME's utility.

While surgery generally benefits patients with MBC [13], the extent of such benefits correlates with individual circumstances and tumor biology. For every 1 mm increase in the size of the patient's tumor [20], the time to 50% mortality after surgery was shortened by 0.16 months. Similar results were found for age [13], number of metastatic sites [40], lymph node status [41], and breast cancer subtypes [18, 20, 42], consistent with previous studies. The effects of metastatic sites and size and breast cancer subtypes on surgery selection were also quantified. The likelihood of surgery selection increases by 15% for patients with ER-positive tumors when isolating this variable.

Despite widespread agreement with prior studies, certain findings from our study have not been thoroughly explored. Some novel findings, such as the reduced effectiveness associated with right laterality, warrant further investigation. A previous study found that patients with right-sided breast cancers were more likely to have locally advanced disease at initial diagnosis [43], which may indirectly influence the surgical outcome. Apart from this, the role of BCS in patients with MBC remains unclear. Our study discovered that the majority of patients with MBC are more appropriately suited for BCS compared with mastectomy, addressing a gap
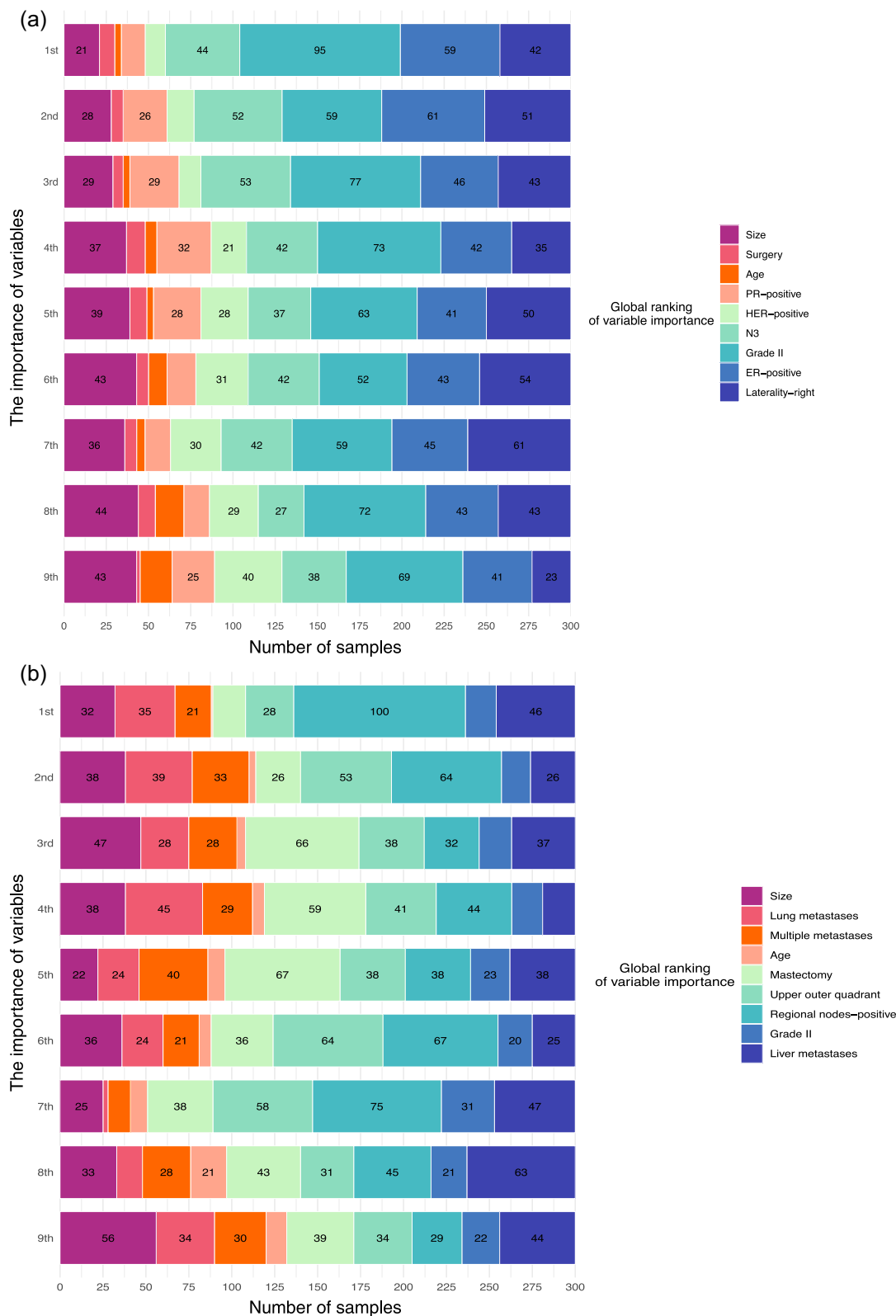
FIGURE 5 Model interpretation based on SurvSHAP(t). (a) Interpretation of deep survival regression with mixture effects of surgery recommendation. (b) Interpretation of deep survival regression with mixture effects of surgery type recommendation. ER, estrogen receptor; HER, human epidermal growth factor receptor-2; PR, progesterone receptor.

previously unexplored in earlier research. We attribute this to observations that patients undergoing BCS typically experience fewer postoperative complications [18] and a reduced psychosocial burden [44], which are factors that could contribute to improved survival rates [45].

The challenge of effective communication among doctors, patients, and families is exacerbated by the absence of tools to visually convey survival benefits. DSME addresses this by offering quantifiable survival advantage data, thereby facilitating clearer discussions and decision-making. Our model appears to be superior in optimizing patient survival compared with traditional treatment guidelines, laying the foundation for the future of precision medicine.

As our research progresses, we aim to refine the DL model, broadening its applicability to a wider range of diseases [46, 47]. Moreover, the development of user-friendly client software for clinical use is anticipated. This software would simplify complex numerical data, making it more accessible for both physicians and patients. With more extensive studies, the DL model could either complement or even potentially replace traditional survival analysis methods, establishing a new standard for personalized treatment recommendations.

The primary limitations of our study stem from the omission of certain prognostic factors. Critical clinical variables, including details of systemic therapy combined with surgery and the timing of systemic therapy initiation, were not included. Incorporating these factors could enhance the precision and focus of model recommendations. Additionally, other outcome metrics, such as progression-free survival, recurrence-free survival, secondary surgery rates, and quality of life, should be further investigated to inform optimal patient treatment strategies. Another limitation is the sourcing of both training and testing datasets from the same database, potentially affecting the models' generalizability. Future studies should aim to validate these models using diverse, real-world clinical data. Nevertheless, this study demonstrates the feasibility of using DL modeling for such applications.

5 | CONCLUSIONS

To the best of our knowledge, this is the first study to use DL models for making individualized surgical recommendations for patients with MBC. DSME exhibited strong performance in both surgery and surgery type recommendations, offering quantitative therapeutic insights. The DL-based therapeutic insights align with current research findings and clinical guidelines, underscoring its potential utility in real-world clinical decision-making.

AUTHOR CONTRIBUTIONS
Enzhao Zhu: Conceptualization (lead); data curation (lead); formal analysis (lead); investigation (equal); methodology (lead); writing—original draft (lead); writing—review and editing (lead). **Linmei Zhang**: Conceptualization (equal); data curation (equal); investigation (equal); methodology (equal); writing—original draft (lead); writing—review and editing (lead). **Jiayi Wang**: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing—original draft (equal). **Chunyu Hu**: Conceptualization (equal); data curation (equal); investigation (equal); methodology (equal); software (equal). **Qi Jing**: Investigation (equal); methodology (equal); validation (equal); visualization (equal). **Weizhong Shi**: Conceptualization (equal); data curation (equal); formal analysis (equal); writing—original draft (equal). **Ziqin Xu**: Data curation (equal); formal analysis (equal); investigation (equal); methodology (equal); writing—original draft (equal). **Pu Ai**: Data curation (equal); formal analysis (equal); funding acquisition (equal); investigation (equal); writing—original draft (equal). **Zhihao Dai**: Conceptualization (equal); data curation (equal); formal analysis (equal); investigation (equal); methodology (equal). **Dan Shan**: Conceptualization (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); visualization (equal); writing—review and editing (equal). **Zisheng Ai**: Funding acquisition (lead); project administration (equal); resources (equal); supervision (equal); validation (equal); visualization (equal); writing—original draft (equal); writing—review and editing (equal).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
This study analyzed public data sets that can be found at the Surveillance, Epidemiology, and End Results Program website (https://seer.cancer.gov/index.html).

ETHICS STATEMENT
The studies involving human participants were approved by the National Cancer Institution (approval ID: SAR0059979).

INFORMED CONSENT

Written informed consent for participation was not required for this study in accordance with national legislation and institutional requirements.

ORCID

Dan Shan http://orcid.org/0000-0002-2762-6770

REFERENCES

1. Andre F, Slimane K, Bachelot T, Dunant A, Namer M, Barrelier A, et al. Breast cancer with synchronous metastases: trends in survival during a 14-year period. J Clin Oncol. 2004; 22(16):3302–8. https://doi.org/10.1200/jco.2004.08.095
2. Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M. Estimation of the number of women living with metastatic breast cancer in the United States. Cancer Epidemiol Biomarkers Prevent. 2017;26(6):809–15. https://doi.org/10.1158/1055-9965.epi-16-0889
3. Bernard-Marty C, Cardoso F, Piccart MJ. Facts and controversies in systemic treatment of metastatic breast cancer. Oncologist. 2004;9(6):617–32. https://doi.org/10.1634/theoncologist.9-6-617
4. Cardoso F, Costa A, Norton L, Senkus E, Aapro M, André F, et al. ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2). Breast. 2014;23(5):489–502. https://doi.org/10.1016/j.breast.2014.08.009
5. Hortobagyi GN. Treatment of breast cancer. N Engl J Med. 1998; 339(14):974–84. https://doi.org/10.1056/nejm199810013391407
6. Gradishar WJ, Moran MS, Abraham J, Abramson V, Aft R, Agnese D, et al. NCCN Guidelines® insights: breast cancer, version 4.2023. J Natl Compr Canc Netw. 2023;21(6):594–608. https://doi.org/10.6004/jnccn.2023.0031
7. Coffey JC, Wang JH, Smith MJF, Bouchier-Hayes D, Cotter TG, Redmond HP. Excisional surgery for cancer cure: therapy at a cost. Lancet Oncol. 2003;4(12):760–8. https://doi.org/10.1016/s1470-2045(03)01282-8
8. Baum M, Demicheli R, Hrushesky W, Retsky M. Does surgery unfavourably perturb the "natural history" of early breast cancer by accelerating the appearance of distant metastases? Eur J Cancer. 2005;41(4):508–15. https://doi.org/10.1016/j.ejca.2004.09.031
9. Marks CE, Thomas SM, Fayanju OM, DiLalla G, Sammons S, Hwang ES, et al. Metastatic breast cancer: who benefits from surgery? Am J Surg. 2022;223(1):81–93. https://doi.org/10.1016/j.amjsurg.2021.07.018
10. Ruiterkamp J, Ernst MF, van de Poll-Franse LV, Bosscha K, Tjan-Heijnen VCG, Voogd AC. Surgical resection of the primary tumour is associated with improved survival in patients with distant metastatic breast cancer at diagnosis. Eur J Surg Oncol. 2009;35(11):1146–51. https://doi.org/10.1016/j.ejso.2009.03.012
11. Budd GT, Cristofanilli M, Ellis MJ, Stopeck A, Borden E, Miller MC, et al. Circulating tumor cells versus imaging: predicting overall survival in metastatic breast cancer. Clin Cancer Res. 2006;12(21):6403–9. https://doi.org/10.1158/1078-0432.ccr-05-1769
12. Danna EA, Sinha P, Gilbert M, Clements VK, Pulaski BA, Ostrand-Rosenberg S. Surgical removal of primary tumor reverses tumor-induced immunosuppression despite the presence of metastatic disease. Cancer Res. 2004;64(6):2205–11. https://doi.org/10.1158/0008-5472.can-03-2646
13. Fields RC, Jeffe DB, Trinkaus K, Zhang Q, Arthur C, Aft R, et al. Surgical resection of the primary tumor is associated with increased long-term survival in patients with stage IV breast cancer after controlling for site of metastasis. Ann Surg Oncol. 2007;14(12):3345–51. https://doi.org/10.1245/s10434-007-9527-0
14. Thomas A, Khan SA, Chrischilles EA, Schroeder MC. Initial surgery and survival in stage IV breast cancer in the United States, 1988–2011. JAMA Surg. 2016;151(5):424–31. https://doi.org/10.1001/jamasurg.2015.4539
15. Gennari A, André F, Barrios CH, Cortés J, de Azambuja E, DeMichele A, et al. ESMO clinical practice guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer. Ann Oncol. 2021;32(12):1475–95. https://doi.org/10.1016/j.annonc.2021.09.019
16. Hu Q, Zhong X, Liu X, Xie Y, Hu K, He P, et al. Resection of primary lesion for patients with metastatic breast cancer: where are we now. Chin Clin Oncol. 2018;7(3):24. https://doi.org/10.21037/cco.2018.05.03
17. Mamtani A, Morrow M. Why are there so many mastectomies in the United States? Annu Rev Med. 2017;68:229–41. https://doi.org/10.1146/annurev-med-043015-075227
18. Babiera GV, Rao R, Feng L, Meric-Bernstam F, Kuerer HM, Singletary SE, et al. Effect of primary tumor extirpation in breast cancer patients who present with stage IV disease and an intact primary tumor. Ann Surg Oncol. 2006;13(6):776–82. https://doi.org/10.1245/aso.2006.03.033
19. Zhu E, Zhang L, Wang J, Hu C, Pan H, Shi W, et al. Deep learning-guided adjuvant chemotherapy selection for elderly patients with breast cancer. Breast Cancer Res Treat. 2024. https://doi.org/10.1007/s10549-023-07237-y
20. Rapiti E, Verkooijen HM, Vlastos G, Fioretta G, Neyroud-Caspar I, Sappino AP, et al. Complete excision of primary breast tumor improves survival of patients with metastatic breast cancer at diagnosis. J Clin Oncol. 2006;24(18):2743–9. https://doi.org/10.1200/jco.2005.04.2226
21. Hankey BF, Ries LA, Edwards BK. The surveillance, epidemiology, and end results program: a national resource. Cancer Epidemiol Biomarkers Prev. 1999;8(12):1117–21. https://pubmed.ncbi.nlm.nih.gov/10613347/
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247–51. https://doi.org/10.1016/j.ypmed.2007.08.012
23. Pan H, Wang J, Shi W, Xu Z, Zhu E. Quantified treatment effect at the individual level is more indicative for personalized radical prostatectomy recommendation: implications for prostate cancer treatment using deep learning. J Cancer Res Clin Oncol. 2024;150(2):67. https://doi.org/10.1007/s00432-023-05602-4
24. Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc Natl Acad Sci USA. 2019;116(10):4156–65. https://doi.org/10.1073/pnas.1804597116

25. Yao L, Chu Z, Li S, Li Y, Gao J, Zhang A. A survey on causal inference. ACM Trans Knowl Discov Data. 15(5):74. https://doi.org/10.1145/3444944

26. Schrod S, Schäfer A, Solbrig S, Lohmayer R, Gronwald W, Oefner PJ, et al. BITES: balanced individual treatment effect for survival data. Bioinformatics. 2022;38(Suppl 1):i60–7. https://doi.org/10.1093/bioinformatics/btac221

27. Li F, Morgan KL, Zaslavsky AM. Balancing covariates via propensity score weighting. J Am Stat Assoc. 2018;113(521):390–400. https://doi.org/10.1080/01621459.2016.1260466

28. Johansson FD, Shalit U, Kallus N, Sontag DA. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. J Mach Learn Res. 2020;23:1–50. https://doi.org/10.48550/arXiv.2001.07426

29. Feydy J, Séjourné T, Vialard F-X, Amari Si, Trouvé A, Peyré G. Interpolating between optimal transport and MMD using Sinkhorn divergences. International conference on artificial intelligence and statistics. 2018. https://doi.org/10.48550/arXiv.1810.08278

30. Nagpal C, Yadlowsky S, Rostamzadeh N, Heller K. Deep Cox mixtures for survival regression. arXiv E Prints. 2021. https://doi.org/10.48550/arXiv.2101.06536

31. Katzman J, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. Deep survival: a deep Cox proportional hazards network. 2018. https://doi.org/10.48550/arXiv.1804.03184

32. Nagpal C, Goswami M, Dufendach KA, Dubrawski AW. Counterfactual phenotyping with censored time-to-events. Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining 2022. 2022. https://doi.org/10.48550/arXiv.2202.11089

33. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J Am Stat Assoc. 2001;96(454):440–8. https://doi.org/10.1198/016214501753168154

34. Krzyziński M, Spytek M, Baniecki H, Biecek P. SurvSHAP(t): time-dependent explanations of machine learning survival models. Knowl Based Syst. 2023;262:110234. https://doi.org/10.1016/j.knosys.2022.110234

35. Lee JS, Toktas O, Soran A. Role of locoregional treatment in de novo stage IV breast cancer. Clin Med Insights Oncol. 2020;14:1179554920942440. https://doi.org/10.1177/1179554920942440

36. Teshome M. Role of operative management in stage IV breast cancer. Surg Clin North Am. 2018;98(4):859–68. https://doi.org/10.1016/j.suc.2018.03.012

37. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134–50. https://doi.org/10.1016/s0140-6736(16)31891-8

38. Huttunen T, Leidenius M, Jahkola T, Mattson J, Suominen S, Meretoja T. Delay in the initiation of adjuvant chemotherapy in patients with breast cancer with mastectomy with or without immediate breast reconstruction. BJS Open. 2022;6(4):zrac096. https://doi.org/10.1093/bjsopen/zrac096

39. Yao K, Sisco M, Bedrosian I. Contralateral prophylactic mastectomy: current perspectives. Int J Womens Health. 2016;8:213–23. https://doi.org/10.2147/ijwh.s82816

40. Blanchard DK, Shetty PB, Hilsenbeck SG, Elledge RM. Association of surgery with improved survival in stage IV breast cancer patients. Ann Surg. 2008;247(5):732–8. https://doi.org/10.1097/sla.0b013e3181656d32

41. Quinn EM, Kealy R, O'Meara S, Whelan M, Ennis R, Malone C, et al. Is there a role for locoregional surgery in stage IV breast cancer? Breast. 2015;24(1):32–7. https://doi.org/10.1016/j.breast.2014.10.009

42. Gnerlich J, Jeffe DB, Deshpande AD, Beers C, Zander C, Margenthaler JA. Surgical removal of the primary tumor increases overall survival in patients with metastatic breast cancer: analysis of the 1988–2003 SEER data. Ann Surg Oncol. 2007;14(8):2187–94. https://doi.org/10.1245/s10434-007-9438-0

43. Al Saad S, Al Shenawi H, Almarabheh A, Al Shenawi N, Mohamed AI, Yaghan R. Is laterality in breast cancer still worth studying? Local experience in Bahrain. BMC Cancer. 2022;22(1):968. https://doi.org/10.1186/s12885-022-10063-y

44. Zehra S, Doyle F, Barry M, Walsh S, Kell MR. Health-related quality of life following breast reconstruction compared to total mastectomy and breast-conserving surgery among breast cancer survivors: a systematic review and meta-analysis. Breast Cancer. 2020;27(4):534–66. https://doi.org/10.1007/s12282-020-01076-1

45. Hofvind S, Holen Å, Aas T, Roman M, Sebuødegård S, Akslen LA. Women treated with breast conserving surgery do better than those with mastectomy independent of detection mode, prognostic and predictive tumor characteristics. Eur J Surg Oncol. 2015;41(10):1417–22. https://doi.org/10.1016/j.ejso.2015.07.002

46. Zhu E, Chen Z, Ai P, Wang J, Zhu M, Xu Z, et al. Analyzing and predicting the risk of death in stroke patients using machine learning. Front Neurol. 2023;14:1096153. https://doi.org/10.3389/fneur.2023.1096153

47. Zhu E, Shi W, Chen Z, Wang J, Ai P, Wang X, et al. Reasoning and causal inference regarding surgical options for patients with low-grade gliomas using machine learning: a SEER-based study. Cancer Med. 2023;12(22):20878–91. https://doi.org/10.1002/cam4.6666

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhu E, Zhang L, Wang J, Hu C, Jing Q, Shi W, et al. Personalized surgical recommendations and quantitative therapeutic insights for patients with metastatic breast cancer: insights from deep learning. Cancer Innov. 2024;3:e119. https://doi.org/10.1002/cai2.119